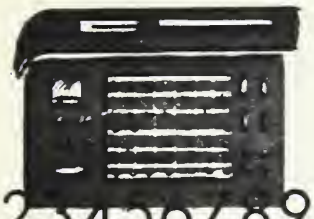
Historic, archived document

Do not assume content reflects current scientific knowledge, policies, or practices.

R31An
no.4

SEP 14 1959
U. S. DEPARTMENT OF AGRICULTURE



MEAN SEPARATION

by the
Functional Analysis
of Variance
and Multiple Comparisons

ARS-20-3 MAY 1957

Agricultural Research Service
UNITED STATES DEPARTMENT OF AGRICULTURE

FOREWORD

This article is written for the investigator who has some understanding of designing experiments and the evaluation of experimental data by the analysis of variance. It discusses (1) the importance of clearly defining the objectives of an experiment before its initiation and emphasizes the need and value of careful planning in research, (2) the application of component sums of squares of the analysis of variance to the biological, chemical, engineering interpretation of an experiment, with particular emphasis on orthogonal and nonorthogonal comparisons, (3) the use of orthogonal polynomials in describing trend relationships, and (4) tests for the separation of means.

Duncan's Multiple Range Test is illustrated and, where appropriate, recommended to the research workers of the Agricultural Research Service.

It is hoped that this article will be helpful in the planning and interpretation of some experiments and that the procedures described will become familiar to those who perform biological, chemical, or engineering research experiments.

Copies of this article may be obtained from Biometrical Services, Agricultural Research Service, Plant Industry Station, Beltsville, Md.

E. L. LeClerg
Chief, Biometrical Services
Agricultural Research Service

CONTENTS

	Page
Preface	i
Introduction	1
Planning of Experiments	1
Testing a Statistical Hypothesis	2
Functional Analysis of Variance	5
Class Comparisons	6
Orthogonal Comparisons	6
Nonorthogonal Comparisons	13
Trend Comparisons	15
Analytical Procedures	15
Multiple Comparison Tests for Separation of Means	20
Least Significant Difference Test (LSD)	21
Duncan's Multiple Range Test	23
Analytical Procedure	24
Table of Significant Studentized Ranges for the 5-percent Level of Significance	25
Table of Significant Studentized Ranges for the 1-percent Level of Significance	26
Data Presentation in Tables	29
Literature Cited	32

MEAN SEPARATION BY THE FUNCTIONAL ANALYSIS OF VARIANCE AND MULTIPLE COMPARISONS

E. L. LeClerg[1/]

INTRODUCTION

Carefully designed experiments are vital to investigative work in all branches of science. The detection and evaluation of differences among particular variables or factors (varieties, breeds, treatments, temperatures, concentrations, density measurements, analytical abilities of a group of persons, time, and others) are assured by properly planning for appropriate comparisons before an experiment is initiated. A little foresight at this time might avoid much disappointment and even failure as far as inductive reasoning from the experimental data is concerned.

In general, an experiment is conducted to answer certain questions posed in advance by the investigator. It should be recognized that the procedure for the analysis of the data is an integral part of designing an experiment. The question or questions to be answered constitute one of the important factors in determining the variables to be included, the design of the experiment, as well as the appropriate method of comparison of the means.

Whenever possible, certain basic principles should be recognized in the selection of the treatment combinations. These principles include: (1) The factorial arrangement when two or more variables are studied, (2) the selection of treatment levels to obtain information on trends or optimum responses, or (3) the use of standards or controls. For these experiments, the analysis and the procedure for evaluation of interrelationships of means is defined. However, the above principles are not applicable to all types of experimentation. In these later situations, multiple comparison tests are required for the separation of means.

PLANNING OF EXPERIMENTS

In planning an experiment, the investigator should have a thorough understanding of the nature of the problem and of any relevant statistical theory associated with it. Frequently this is not so easy as it sounds. It is necessary to state not only what is to be tested but also to indicate the inferences to be made from the results. Thus, it

---

[1/]Chief, Biometrical Services, Agricultural Research Service, U. S. Department of Agriculture, Plant Industry Station, Beltsville, Md.

The writer expresses appreciation to the professional staff of Biometrical Services for their help and suggestions in the preparation of the manuscript.

cannot be overemphasized that the time to consider statistical inferences is at the time the experiment is being planned. Unfortunately, consideration of statistical analysis is too frequently deferred until the experiment is completed, when, as a matter of fact, the underlying principles of statistics should have been applied in planning the experiment.

It is good policy, before conducting an experiment, to prepare a written statement including (1) the objectives; (2) the description of the experiment, including the design, experimental variables, number of replications, sampling procedure, etc.; and (3) the method of analysis of the results.

It should be pointed out that, in deciding upon an experimental design, the statistical considerations while very important are never paramount. Statistical procedures have been emphasized so much in recent years that there is a real danger of statistics being regarded as the important part of experimentation. A design must be operationally practical whether it be for field, animal, or laboratory experimentation.

In the application of the principles of statistics to experiments, there is no clear line of demarcation between design and analysis. The design chosen predetermines, to a large degree, the analysis of the data. It is not within the scope of this discussion to develop the various designs applicable to field, greenhouse, animal, or laboratory experimentation. Instead, the discussion will be limited to the testing of significance of means by the functional analysis of variance and by certain multiple-comparison procedures.

Before a discussion of the procedures for the evaluation of an array of means is given, consideration of the testing of statistical hypotheses is desirable.

TESTING A STATISTICAL HYPOTHESIS

The analysis of variance is a statistical procedure to aid in interpreting data, but is not itself a summary of the experiment. Proper interpretation will involve an examination of the means in order that appropriate inferences may be made in terms of the particular research discipline and not merely in statistical jargon. Before discussing the evaluation and separation of means, it is desirable that the meaning of testing a statistical hypothesis be understood and its application appreciated.

Because of the experimental error associated with each mean of an experiment one never proves an hypothesis. All that can be inferred is that, on the basis of the information supplied by observation of a

sample and within certain limits of probability, the observed mean represents the population value. This is an important fact to recognize when attempting to make statistical inferences.

It is also important to recognize that by statistical methods one can come nearer to proving that something is not true about a population than that something is true. Thus, through this negativistic approach a hypothesis is not tested directly. Instead, an opposite hypothesis is formulated, called the null hypothesis, which is tested on the basis of the evidence presented by the sample. Specifically, the null hypothesis assumes that the set of observed differences between means could be a random set from one population with a mean of zero. Like most scientific procedures, this negativistic approach is a cautious way of proceeding.

A test of significance is a procedure to determine whether some quantity, which is subject to random fluctuation, differs from a theoretical value by an amount greater than that explained by random variation alone. In making tests of significance, the researcher must decide upon the level of significance appropriate for the inferences to be drawn from the data. Such a level of significance (for example, the 5-percent level) indicates the chances of erroneously concluding a departure from the null hypothesis.

Since there is fluctuation in sample means and sample standard deviation, the decision as to whether the null hypothesis shall be accepted or rejected will be based on (1) the information obtained from the experimental data and (2) the risk one is willing to take that the decision may be wrong. The errors of rejecting or not rejecting a null hypothesis are classified into two groups. An investigator may decide to

1) reject the null hypothesis, when in fact, though unknown to the investigator, it is actually true. In other words, one asserts that a real difference exists when it does not, and it is concluded that a sample differs from the others when in reality all the discrepancy is due to chance causes. The assertion of a significant difference when none exists is called an error of the first kind or false rejection (Type I).

2) not reject the null hypothesis or, in other words to accept the null hypothesis, when in fact, it is wrong. In this instance, one asserts that a real difference does not exist when one really does exist. Or, one does not detect the presence of some assignable cause of either poor or better performance. This assertion of a nonsignificant difference when one exists is called an error of the second kind or false acceptance (Type II).

The likelihood of committing a Type I error is under the control of the investigator and is determined by the level of significance specified for testing the null hypothesis. As the level of significance decreases (i.e., for example, from the 1-percent to the 5-percent level), more errors of Type I are likely to be committed. Conversely, the higher the level of significance chosen (better Type I protection), the more likely are Type II errors to be committed. Thus, at the 5-percent level of significance, one would expect to reject the null hypothesis not more than 5 out of every 100 times when in fact the hypothesis is actually true. The probability of committing a Type II error, however, cannot be stated precisely unless an alternative hypothesis can be specified. But it follows that if very high levels of significance are set, differences between samples from separate populations may not be judged to be significant. The frequency with which Type II errors occur depends upon several factors, chief of which is the precision with which the statistic is calculated.

The custom of using the 5-percent level of significance has led many experimenters to place complete confidence in this level of risk protection for controlling Type I errors. Whether or not this is the most appropriate level depends upon the nature of the experiment, the possible economic relationship from the research, and the number of replications used. In the preliminary stages of an investigational program, the experimenter may be investigating a number of treatments or varieties to determine the advisability of further experimentation. At this point in the program material may be available for only a limited number of replications. In such a situation it may be deemed advisable to use a 10-percent level of significance test in order to have a Type II error (error of failing to detect a specified difference when the treatments actually differ by more than such an amount) of reasonable size. If in the more advanced stage of a program, the investigator wishes to use the 1-percent level of significance test instead of the 5-percent level, the number of replications or samples must be increased in order to more adequately protect against a Type II error.

The action to be taken regarding these two types of errors depends to a large degree on the relative seriousness of committing the errors under a specified set of conditions and the inferences to be drawn from the results of the experiment. It is apparent, of course, that a design and the number of replications should be chosen which minimizes the probabilities of both Type I and Type II errors. Accordingly, a level of significance must be chosen with regard to the relative importance of the risk involved in committing either of these errors.

Increasing the precision of an experiment decreases the risk of both types of errors, but particularly those of the second kind. One method of accomplishing this is to increase the number of replications.

Cochran and Cox (2) discuss this problem and give a table for determining the number of replications required for a given probability of obtaining a significant result.

No investigator wants to be shown by further work that his conclusions were erroneous. Therefore, he will want to keep low the risk of concluding that a certain difference is real when in fact it is not. At the same time, it is important not to miss positive effects that are real.

Thus, consideration of the exact nature of the Type I and Type II errors is very important in drawing inferences from data. When experimental data are highly variable and comprise only a few observations, the test of significance controls only the error of the first kind at the arbitrary level of risk. Under such circumstances the investigator would regard as unimportant all but very large differences. Thus, because of low sensitivity, a real and possibly important difference may well have occurred.

The investigator in designing his experiment must choose a sufficient number of observations or replications so as to maximize the possibilities of picking out, by means of a test of significance, the sufficiently large differences which may be of practical importance. Planning experiments in this way controls the risk of failing to detect a real departure from the null hypothesis. Present experimental designs control errors of the first kind (Type I) satisfactorily, but without proper attention to errors of the second kind (Type II) advancement will be impaired and its costs will exceed what is required to achieve the desired results.

FUNCTIONAL ANALYSIS OF VARIANCE

The functional analysis of variance refers to testing the significance of the individual or group comparisons comprising a number of treatment effects in an experiment. These comparisons can be tested by segregating the degrees of freedom for treatment effects into meaningful single degrees of freedom and associated sums of squares. As a consequence, it follows that each degree of freedom for treatment effects in an analysis of variance can provide potential information of some biological or chemical significance.

It is worthy of note that the responsibility for deciding upon the appropriate comparisons is largely that of the experimenter, since they are dependent upon the question or questions to be answered from the experiment.

For purposes of this paper two types of comparisons will be

discussed. These are:

1. Class comparisons

2. Trend comparisons

Class Comparisons

Class comparisons comprise different types of variables assignable to classes according to their nature or the effect they produce. The comparisons to be tested for significance should be planned as part of the experiment; otherwise, the comparisons may only indicate peculiarities of the particular data and not the true relationships desired from the experiment.

Where two or more sets of comparisons are to be compared simultaneously, it is desirable that they be independent of each other (i.e., orthogonal) if the corresponding components in the analysis of variance are also to be independent. It is possible to partition a sum of squares into many different orthogonal sets. In any given experiment, ordinarily only one set would be selected at the planning stage.

Nonindependent or nonorthogonal comparisons tend to be affected by each other, and the separation of their corresponding components in the analysis results in ambiguous probability statements relating to the interpretation of the data.

The remainder of this discussion of class comparisons will be limited to that relating to orthogonal and nonorthogonal comparisons.

Orthogonal Comparisons

For illustrative purposes it is assumed that an investigator is interested in comparing five treatments, such as fertilizers, concentrations, and temperatures. The four comparisons tabulated below are referred to as a set of orthogonal comparisons:

Comparison	Treatments				
	1	2	3	4	5
C_1	-1	+4	-1	-1	-1
C_2	+1	0	+1	-1	-1
C_3	+1	0	-1	0	0
C_4	0	0	0	+1	-1

The numbers indicated in the above table are termed coefficients and the plus or minus signs indicate the nature of the comparisons. Thus, in the above comparisons

C_1 = Treatment 2 vs. average of other treatments.
C_2 = Treatments 1 and 3 vs. treatments 4 and 5.
C_3 = Treatment 1 vs. treatment 3.
C_4 = Treatment 4 vs. treatment 5.

Two important rules relate to the coefficients in the table, namely, that the sum of the coefficients of a comparison must equal zero and that the sum of the products of the corresponding coefficients of any two comparisons must be zero. When both of these conditions are satisfied, the comparisons are independent of one another and are said to be "orthogonal".

It is apparent from the table that the sum of the coefficients of each comparisons equal zero. Thus, for the first comparison

$$C_1 = -1 + 4 - 1 - 1 - 1 = 0$$

Likewise, the sum of the products of the coefficients of any two comparisons is zero. For example:

$$C_1 \times C_2 = (-1)(+1)+(+4)(0)+(-1)(+1)+(-1)(-1)+(-1)(-1) = 0$$

Thus, it appears that the conditions of orthogonality are satisfied.

The following breakdown of the four degrees of freedom associated with these five treatments would correspond to the four orthogonal comparisons presented in the above table:

Source of variation	Degrees of freedom
Treatments	4
Treatment 2 vs. others	1
Treatment 1 and 3 vs. treatments 4 and 5	1
Treatment 1 vs. treatment 3	1
Treatment 4 vs. treatment 5	1

When the number of orthogonal comparisons is equal to the degrees of freedom between treatments, the sum of squares computed with the coefficients in each row will account for the variability between treatments as measured by the sum of squares for treatments in the basic analysis of variance.

Several examples, in different areas of research, will be used to illustrate the procedure of determining significance of class comparisons. It is hoped that these examples may serve as a practical guide to the research worker in the interpretation of experimental data.

Example 1 (Soils and Fertilizers)

An experiment is planned to study the effect of different sources of nitrogen fertilizer on the yield of a crop. The sources of nitrogen will be ammonium sulfate, ammonium nitrate, and urea. A nonfertilized control is to be included. The experiment will comprise five replications. Since the sum of squares for sources of nitrogen comprise the sum of squares for the individual sources of nitrogen, it is possible to break down the degrees of freedom in the following manner:

Basic Form	D.F.	Functional Form	D.F.
Replications	4	Replications	4
Sources[1]	3	Sources[1]	3
Error	12	(a) Nitrogen vs. no nitrogen	1
		(b) Organic vs. inorganic	1
		(c) Amm. sulfate vs. amm. nitrate	1
		Error	12

[1] Includes the control

It seems desirable at this point to discuss the arithmetic procedure required in calculating the sums of squares for the three individual degrees of freedom associated with the sources of nitrogen. The hypothetical totals (sum of the yields of all plots for individual nitrogen sources) for each of the sources of nitrogen and control required for this type of analysis are:

Nitrogen sources	Total yield
None	1092
Urea	1226
Ammonium sulfate	1155
Ammonium nitrate	1176

A scheme for separating the three degrees of freedom into orthogonal comparisons for sources of nitrogen in this sample is as indicated below:

Comparison	Sources of nitrogen			
	None	Urea	Ammonium sulfate	Ammonium nitrate
C_1 = [Nitrogen] vs. [No nitrogen]	+3	-1	-1	-1
C_2 = [Organic] vs. [Inorganic nitrogen]	0	+2	-1	-1
C_3 = [Amm. sulfate] vs. [Amm. nitrate]	0	0	+1	-1

The means of organizing the necessary calculations to determine the sum of squares for each comparison can be accomplished by either of two procedures. Each of these will be discussed.

Procedure (1)

The above table of coefficients gives the signs appropriate for combining the observations to obtain estimates of the effects of the three comparisons on yield. The contribution of each effect to the total sum of squares for sources of nitrogen is obtained by squaring the effect estimates and dividing by the sum of the squares of the coefficients involved in the comparison. These sums of squares must be on a unit plot basis so that the number of replications in this case, five, must also appear as a part of the division.

Therefore, the sum of squares for the comparison of "nitrogen vs. no nitrogen" (C_1) is

$$C_1 = \frac{[3(\underset{\text{None}}{1092}) - 1(\underset{\text{Urea}}{1226}) - 1(\underset{\text{Ammonium sulfate}}{1155}) - 1(\underset{\text{Ammonium nitrate}}{1176})]^2}{[(3)^2+(-1)^2+(-1)^2+(-1)^2]\ (5)}$$

There is an easy rule for determining the entries in the denoninator. First, the coefficients of the comparison under consideration are written down, as

$$C_1 = +3\ -1\ -1\ -1$$

Then, as indicated above, the coefficients are squared and the squares summated, thus

$$(3)^2+(-1)^2+(-1)^2+(-1)^2 = 12$$

This value, 12, is then multiplied by the number of replications, which in this case was 5. Proceeding with the determination of the sum of squares for the C_1 comparison, we get

$$C_1 = \frac{(3276-1226-1155-1176)^2}{(9+1+1+1)\ (5)}$$

$$= \frac{(3276-3557)^2}{60} = 1{,}316.02$$

Procedure (2)

The sum of squares for orthogonal comparisons can be calculated by

another procedure. This involves the setting up of the following table:

Comparison	Treatments and totals				Diff.	$(\text{Diff.})^2$	Divisor	Sum of Squares
	None 1092	Urea 1226	Ammonium sulfate 1155	Ammonium nitrate 1176				
C_1	+3	-1	-1	-1	-281	78,961	60	1316.02
C_2	0	+2	-1	-1	121	14,641	30	488.03
C_3	0	0	+1	-1	- 21	441	10	44.10

The value for the "difference" for the C_1 comparison, for example, is given by

$$+3(1092)-1(1226)-1(1155)-1(1176) = -281$$

The sum of squares is then computed as

$$\frac{(-281)^2}{12x5} = \frac{78,961}{60} = 1316.02$$

which is the same value as determined by Procedure (1). It should be pointed out that the divisors are the same as determined for the denominator of Procedure (1).

The sums of squares for the three individual degrees of freedom indicated above should, in an experiment with orthogonal comparisons, add up to the total sum of squares corresponding to the three degrees of freedom for "sources of nitrogen." The mean square for each single degree of freedom can be tested for significance by means of the F-test.

From the above discussion, it is apparent that a notable increase in information is furnished from an experiment designed to evaluate individual comparisons among a set of treatments.

Example 2 (Chemical)

A slightly different method of splitting up the degrees of freedom is useful in factorial experiments. Suppose that A and B represent two concentrations of two different chemicals, then the following three comparisons are obvious:

Comparison	Chemical I		Chemical II	
	Conc. A	Conc. B	Conc. A	Conc. B
C_1 = Between chemicals	-1	-1	+1	+1
C_2 = Between concentrations	-1	+1	-1	+1
C_3 = Interaction	+1	-1	-1	+1

The third comparison (C_3) represents the extent to which the

chemicals give different results at the two levels and is comparable to an interaction between chemicals and concentrations.

Example 3 (Plant Pathology)

An investigator might be studying the protective action of six fungicides in five replications, three of the fungicides being copper compounds and three sulfur. The five degrees of freedom for fungicides can be separated into one degree of freedom associated with the difference between groups of fungicides (i.e., copper and sulfur) and four degrees of freedom associated with the differences within the two groups of compounds. These four degrees of freedom can be further subdivided into two degrees of freedom each for copper and sulfur fungicides, respectively. The degrees of freedom would be as follows:

Basic Form	D.F.
Replications	4
Fungicides	5
Error	20

Functional Form	D.F.
Replications	4
Fungicides	5
(a) Between groups	1
(b) Within groups	4
(i) Between copper cmpds.	2
(ii) Between sulfur cmpds.	2
Error	20

Thus, the F-test for the mean square for the one degree of freedom for "between groups" indicates immediately whether the protective action of copper and sulfur, as groups, is the same or different under the conditions of the experiment. Likewise, F-tests for the copper and sulfur compounds indicate whether significant differences exist among the three compounds comprising each group. These comparisons could be easily missed if the five degrees of freedom between fungicides were tested together as indicated in the basic form of the analysis of variance.

Example 4 (Plant Pathology)

A greenhouse experiment might be made to determine the relative pathogenicity of three isolates of Rhizoctonia solani as the cause of damping-off of sugar beets. Each of the five replications may consist of five pots of each isolate and a noninoculated control. The analyses of variances are, as follows:

Basic Form	D.F.
Replications	4
Isolates[1/]	3
Error	12

Functional Form	D.F.
Replications	4
Isolates[1/]	3
(a) Between isolates	2
(b) Isolates vs. no isolates	1
Error	12

1/ Includes the control.

This functional analysis permits an F-test as to (1) whether there are significant differences among the three isolates as related to pathogenicity or (2) whether the stand of sugar beets was significantly greater for the average of the three isolates as compared with the noninoculated control.

One needs to be careful to use an appropriate error term for making the F-tests in experiments in which a control (nontreated) is a feature. It will be apparent that certain treatments are likely to be more variable in their effects than others. As a consequence of this differential variability, comparisons comprising treatments with erratic effects will have a higher experimental error than those with more stable treatments.

This situation may be important because one of the assumptions of the analysis of variance is that the experimental errors for difference treatments must have a common variance; in other words, the errors should represent the same population values. Under conditions of differential variability of individual treatments this assumption may not be adequately fulfilled.

In such cases it is sometimes helpful to calculate a separate error term for any specific comparison. Judgment, however, must be exercised in using this procedure as it should be recognized that some of the partitioned error terms might be based on a small number of degrees of freedom. It may be advisable to seek assistance from the biometrician on problems of this type.

Example 5 (Chemistry)

Let it be desired to determine the concentration of iron in a solution by 10 different analysts. Each of five female and five male analysts make five determinations of the iron solution. By breaking down the nine degrees of freedon for the 10 analysts, one can get an F-test directly to determine whether any difference exists between the male and female analysts. Furthermore, the eight degrees of freedom for "within groups of analysts" can be subdivided to permit a test of significance (F-test) for differences between male analysts and between female analysts. This is shown below:

Basic Form	D.F.
Between analysts	9
Between samples	40

Functional Form	D.F.	
Between analysts	9	
(a) Male vs. female	1	
(b) Within groups of analysts	8	
(i) Between males		4
(ii) Between females		4
Between samples	40	

Example 6 (Industrial)

It may be desired to test the abrasive action of 4 types of washing machines on a certain kind of wearing apparel. In such an experiment,

replications are represented by the different days in which the test is repeated with the group of four machines. The four machines are designated as A, B, C, and D. Machine A represents the current model whereas the other 3 machines are experimental models developed to replace machine A.

The 3 degrees of freedom for machines can be broken down into the following 3 selected orthogonal comparisons and an F-test performed for each single degree of freedom:

Comparison	Nature of Comparison	Machine			
		A	B	C	D
C_1	A vs. others	+3	-1	-1	-1
C_2	B and C vs. D	0	+1	+1	-2
C_3	B vs. C	0	-1	+1	0

It will be noted that the above comparisons form an orthogonal set since the coefficients follow the rule that their sums and sums of products must be zero.

Nonorthogonal Comparisons

Not infrequently it is desirable to design an experiment in order to make certain comparisons that are not orthogonal to one another. It should be emphasized that in such situations the probability statements accompanying the test of significance (for example, the 5-percent level) are not accurately discernible. Under such situations care should be exercised in interpreting the experimental data.

Because of the nonorthogonal nature of this type of analysis, the total of the sums of squares for the individual treatment comparisons will deviate from the combined sums of squares for treatments in the basic form of the analysis of variance.

Two examples of experiments comprising nonorthogonal comparisons are given.

Example 1 (Nutrition)

A simple experiment of this type might comprise the daily gains of sheep fed 4 different rations. The first ration might be a standard protein ration which is to be compared with the three others.

The comparison (nonorthogonal) of each ration with the control is symbolically represented by the following coefficients:

Comparison	Rations			
	1(Control)	2	3	4
C_1 = 1 vs. 2	+1	-1	0	0
C_2 = 1 vs. 3	+1	0	-1	0
C_3 = 1 vs. 4	+1	0	0	-1

From the above table it is apparent that the sum of the coefficients in any comparison is zero, but the sum of the products in each pair of lines is greater than zero. For example, the sum of the products of the coefficients of the first two comparisons is

$$C_1 \times C_2 = (+1)(+1)+(-1)(0)+(0)(-1)+(0)(0) = 1$$

Thus, the above comparisons are not independent, therefore nonorthogonal, but they do give the answers to the questions for which the experiment was designed.

Example 2 (Soil Bacteriology)

At times it may be desirable to be interested in certain nonorthogonal comparisons in experiments dealing with soil amendments. Such might be the case with the evolution of carbon dioxide from fungi in soil to which certain additives were incorporated into the substratum upon which they grow.

The additives might include:

1. Control (no additive)
2. Alfalfa meal
3. Corncob meal
4. Corncob meal + Mg SO_4

The experiment can be planned to compare each organic additive (alfalfa meal and corncob meal) and chemical treatment with others similar in all respects except one. The three comparisons might be:

Comparison	Control	Alfalfa	Corncob	Corncob +Mg SO_4
C_1 = Alfalfa meal	+1	-1	0	0
C_2 = Meals	0	+1	-1	0
C_3 = Mg SO_4	0	0	-1	+1

While the above comparisons are not orthogonal they do furnish answers having practical significance.

Trend Comparisons

Analytical Procedures

By now it should be obvious that the procedure of splitting up the degrees of freedom and the corresponding sums of squares may have a definite logical basis. Frequently an investigator might be studying a variable at several levels, such as levels of nitrogen, several temperatures, concentrations of a chemical, or levels of a nutrient in a ration. In such cases he is interested in the nature of the trend to the varying levels of the treatment and the analysis should be designed to evaluate this trend. This would involve the relation between the independent variable (such as nitrogen, temperature, and concentration of chemical) and the dependent variable levels to determine whether the relationship is linear (straight line) or whether there is appreciable curvature. Obviously, in experiments of this type it becomes more informative to know whether the relation between the independent variable and the dependent variable can be regarded as linear or curvilinear rather than test for significance between the means of the various levels of the dependent variable.

Such a situation is found in the data published by Larmour (11). He gave the reaction of flours from 17 varieties of wheat baked with four levels, increasing in equally spaced intervals, of potassium bromate. As an average of the 17 varieties, the totals for loaf volumes for each of the four levels of bromate are given in Table 1.

Table 1. Total loaf volumes for 17 varieties of wheat baked with four levels of potassium bromate

Item	Levels of bromate (mg.)				Grand total
	1	2	3	4	
Loaf volume	154.5	154.7	147.9	141.9	599.0

This example presents a different way of splitting up the degrees of freedom for treatment effects (in this case, levels of bromate). Before proceeding with the method of analysis of such data, it is important to remember that wherever possible the levels of the independent variable (in this example, levels of bromate) should be in equally spaced steps or intervals. This equal spacing of variables also applies to the natural scale, logarithmic scale, or some other simple transformation. Examples of equal spacing on the natural and logarithmic scales are, as follows:

Scale	Levels of Variable
Natural	0, 2, 4, 6, 8, 10, etc
Logarithmic	1, 2, 4, 8, 16, 32, etc

Thus, in the above example the level of potassium bromate increased by concentrations of 1 milligram. When such a procedure is followed, economy in computational work can be attained by the use of orthogonal polynomials. This is especially true if the trend desired is of higher degree than the second-degree polynomial (quadratic).

At this point, it is appropriate to explain and describe the method of regression analysis by means of orthogonal polynomials. If the independent variable is equally spaced, such as in time or space, the convenient method of curve fitting by orthogonal polynomials can be used. The advantage over the usual regression methods of fitting nonorthogonal polynomials arises from the fact that orthogonal polynomials are so constructed that any term of the polynomial is independent of any other term. This property of independence permits the computation of each regression coefficient independently of the others and also facilitates testing the significance of each coefficient.

The summation method, developed by Fisher (8) is a convenient method of fitting orthogonal polynomials. In practice, however, the polynomial method ξ' presented by Fisher and Yates (9) is generally more expeditious if tests of significance of the succeeding terms of the polynomial are desired. The utility of this method depends upon the availability of tables of ξ' values which have been published by Anderson and Houseman (1) and Fisher and Yates (9). The use of the ξ' method will be discussed in detail.

The necessary conditions for orthogonality with respect to polynomials are the same as those previously discussed for the case of orthogonal comparisons. These conditions are illustrated by the orthogonal polynomial values (ξ) in Table 2. For example, the sum of the values of ξ'_3 for n = 5 equals zero and the sum of products of the values of ξ_3 and ξ_4 for n = 5 is equal to zero. Thus

$$-1 + 2 + 0 - 2 + 1 = 0$$
$$(-1)(+1)+(+2)(-4)+(0)(+6)+(-2)(-4)+(+1)(+1) = 0$$

Table 2. Partial table of polynomial (ξ') coefficients

Number of Observations (n) and degrees of polynomial (ξ')									
n=3			n=4			n = 5			
ξ'_1	ξ'_2	ξ'_1	ξ'_2	ξ'_3	ξ'_1	ξ'_2	ξ'_3	ξ'_4
-1	+1	-3	+1	-1	-2	+2	-1	+1
0	-2	-1	-1	+3	-1	-1	+2	-4
+1	+1	+1	-1	-3	0	-2	0	+6
		+3	+1	+1	+1	-1	-2	-4
					+2	+2	+1	+1
Divisor 2	6	20	4	20	10	14	10	70

The analysis of variance for regression components of the above example differs from a simple analysis in that the treatment sum of squares are divided into linear, quadratic, and cubic components.

This general procedure followed in such a supplementary analysis results in the same information as derived from the ordinary regression analysis, but computationally is simpler. The procedure makes possible the subdivision of the degrees of freedom and effects into the appropriate corresponding sums of squares for treatment regression components. Thus, in this example, there are three degrees of freedom for levels of bromate (treatments) that are divided into linear, quadratic, and cubic regression components. The sum of squares and degrees of freedom are correspondently divided for these supplementary tests of significance. To fit a straight line (linear) requires the values of ξ'_1 only, the quadratic requires ξ'_1 and ξ'_2, the cubic requires ξ_1, ξ'_2, ξ'_3, and so forth.

The use of orthogonal polynomials will now be applied to the set of data in Table 1. In the first place, the usual analysis of variance must be made of the data. This computation gives the following:

Source	D.F.	Sum of Squares	Mean Squares
Varieties	16	70.01	4.38
Treatments	3	6.59	2.20
Error	48	5.23	0.11

It is apparent that the value of F = 20.0 (F = 2.20/0.11 = 20.0) for treatments is highly significant and thus the null hypothesis that the four levels of bromate come from the same population is rejected. The important aspect of such a problem, however, is to determine the trend relationship of the data. The three degrees of freedom for treatments, therefore, should be subdivided into linear, quadratic, and cubic components.

The calculations for determining the sums of squares for the three regression components are derived from the treatment totals of Table 1 in the following manner:

Linear Component (ξ'_1)

$$= \frac{[\,-3(154.5) -1\,(154.7) +1\,(147.9) +3\,(141.9)\,]^2}{20 \times 17} = 5.85$$

Quadratic Component (ξ'_2)

$$= \frac{[\,1(154.5) -1\,(154.7) -1\,(147.9) +1\,(141.9)\,]^2}{4 \times 17} = 0.56$$

Cubic Component (ξ_3')

$$= \frac{[\ -1(154.5) +3\ (154.7)\ -3\ (147.9) +1\ (141.9)\ ]^2}{20 \times 17} = 0.18$$

It should be noted that the coefficients used in the above equations come from those given in Table 2. For polynomial values for higher-degree polynomials and greater numbers of observations reference is made to Anderson and Houseman (1) or Fisher and Yates (9). The denominators of the above equations represent the respective divisors of Table 2 and the number of varieties, 17, in this experiment. The divisors represent the sum of the squares of the coefficients for a regression component.

The error term (0.11) in the analysis of variance (see below) is appropriate for testing the significance of the mean squares for treatments. Likewise, this error term can be used to test the significance of the three regression components of the treatment effect (levels). The analysis of variance for treatments now becomes:

Source	D.F.	Sum of Squares	Mean Squares	F
Treatments (levels)	3	6.59		
(a) linear	1	5.85	5.85	53.2**
(b) quadratic	1	0.56	0.56	5.1*
(c) cubic	1	0.18	0.18	1.6
Error	48	5.23	0.11	

The F-tests indicate that the differences between the treatments are chiefly linear effects ($F = \frac{5.85}{0.11} = 53.2$). This indicates that, on the average, there is a tendency for loaf volume to be reduced proportionately with added quantities of potassium bromate. It is apparent, however, that the quadratic regression term is also significant ($F = \frac{0.56}{0.11} = 5.1$). The evidence for such a relationship is apparent from an examination of the totals in Table 1. Thus, it is noted that there is a slight increase in loaf volume from level one to level two but a distinct drop for the third and fourth levels.

It should be clear by now that the analysis of trends has furnished information of value which could not have been obtained by tests of significance for the treatment means (or totals).

Some examples of experiments, in the various fields of agricultural research, where trend (regression) analysis would be appropriate are discussed below.

Examples

(1) Soil fertility experiments involving different rates of application of elements can be set up so that the use of the orthogonal-polynomial method indicates trend relationships. The point at which the yield curve breaks and starts downward would represent the optimum rate of application. With two or more factors in a factorial experiment the trend relationship of the main effects, as well as the interactions, can be evaluated.

(2) Regression analysis would be appropriate in feeding experiments with insects where certain morphological measurements are taken at equal intervals of time.

(3) Many nutritional experiments with large and small animals lend themselves to trend analysis. For instance, a study of different proportions of cottonseed meal in the feed for steers might comprise rations of alfalfa meal plus varying amounts of the cottonseed meal. The rations might contain equally spaced amounts, as 40, 30. 20, and 10 percent cottonseed meal. A regression analysis of weight gain would indicate the optimum amount of cottonseed meal for maximum growth. A control of no cottonseed meal is not necessary unless information is not available as to whether any of the amounts of cottonseed meal would have an effect on gain.

(4) Experiments to determine the age at which a perennial crop, such as asparagus, begins to decrease in yielding ability are well suited to regression analysis. In this case, the seasons become the fixed variable and are equally spaced so that the use of orthogonal polynomials becomes a convenient method of curve fitting and testing.

(5) Regression analysis is appropriate for experiments designed to study the effect of aggregate size on compaction of certain heavy soils at different applied pressures. In an experiment of this type, the various levels of applied pressure should be selected at equally spaced intervals in order to simplify the statistical computations.

(6) The study of spray losses in sprinkler irrigation involves many factors, some of which are best approached by regression analysis. An investigation of the relationship of distance from the sprinkler head to the size of water droplets is well suited for the application of orthogonal polynomials when the distances under study are equally spaced.

(7) Investigations carried out to determine the fungicidal action of certain antibiotic compounds and their ability to penetrate into crop seeds can be analyzed by regression analysis. In such experiments, one might study the effect of increasing concentrations (in equal intervals) of the antibiotic on the percentage of seeds developing a specific plant-pathogenic fungus.

(8) In order to study the effects of sex hormones on small animals with relation to average growth and cumulative food consumption, it is advisable to measure the trend responses of these two variables at intervals of equally spaced days.

(9) Equal spacing of hours would be appropriate in studying the influence of length of photoperiod (in hours) on the number of days required from emergence to flowering of a crop variety.

MULTIPLE COMPARISON TESTS FOR SEPARATION OF MEANS

As pointed out previously, the analysis of variance (F-test) indicates whether the factor means differ significantly as a group, that is, whether they may be considered as representatives of the same population (F value below the 5-percent or 1-percent level of significance) or from different populations (F value above the 5-percent or 1-percent level of significance). However, an indication of nonsignificance does not imply that true effects do not exist. Actually it merely implies that an observed apparent effect would arise more frequently than 1 in 20 (or 1 in 100) times by chance if no real effects existed.

A significant value of F, however, does not indicate which mean might differ significantly from some other mean. But in some experiments, there is no meaningful reason for partitioning the degrees of freedom, as previously discussed, into class or trend comparisons. In such situations, the investigator, may, and often does, wish to know which means may be considered as different from the others. This is probably one of the most frequent decisions confronting an experimenter.

The method for discriminating between means is commonly known as multiple comparison procedure and has received a great deal of attention from mathematical statisticians in the past several years. The important aspect of this problem is that the experimenter achieve his purpose with a specified probability that all the comparisons made will lead to correct answers.

A large number of procedures have been developed for testing the difference among several means. These include the following:

Multiple Range Tests:

1) Least-significant-difference test as a multiple t test (7),
2) Newman-Keuls (10, 12),
3) Tukey's test based on allowances (14),
4) Tukey's X-procedure (15),
5) Duncan's multiple range test (5).

Multiple F Tests:

1) Scheffe's test based on contrasts (13)
2) Duncan's multiple comparisons test (3,4).

These tests vary considerably, and it is difficult for the experimenter to decide upon which one to choose for any given problem. The range tests are applicable to pairwise comparisons of means, whereas the multiple F tests are of a more general nature and are more appropriate for contrast comparisons. Where the experimenter is concerned with testing certain contrasts involving more than two means a multiple F-test is appropriate. However, the amount of work involved in applying a multiple F-test is many times greater than that for the range test.

In biological and chemical research the emphasis is usually on pairwise comparisons. Therefore, range tests are more appropriate. In choosing between the several range tests, the experimenter is confronted with the relative importance of and protection of Type I and Type II errors afforded by the tests. In this regard, Duncan's Multiple Range Test assures the best Type II error protection. This type of protection, in most cases, is the one desired by the investigator. Some of these range tests are no more than mentioned in the following discussion, whereas others will be discussed in considerable detail.

A detailed discussion of multiple comparison procedures is given by Federer (6).

Least Significant Difference Test (LSD)

Fisher (7) first proposed the use of the often used least-significant difference (LSD) test in 1935. Subsequently, it has been discussed in many textbooks on statistics. In this test, the difference between two means is declared significant, at any desired level of significance, if it exceeds the value derived from

$$LSD = t\sqrt{2}\ s_{\bar{x}}$$

where

t = tabulated value of "t" at any desired level (usually the 5-percent level)

$s_{\bar{x}}$ = standard error of the mean.

In other words, the LSD test utilizes the standard error of a difference between two means, $\sqrt{2}\ s_{\bar{x}}$, which serves as the least significant difference between two means when multiplied by the tabulated value of "t" at the desired level of significance. This test is to be applied only when the F test for the homogeneity of the means in the experiment is significant, i.e., that the array of means are not representative of the same population.

Very frequently, especially in animal data, the LSD test is applied to the difference between means which are estimated from

unequal number of observations. In this case

$$LSD = t\sqrt{s^2 \frac{n_1 + n_2}{n_1 n_2}}$$

Most textbooks on statistical methods have tended to evade discussing multiple comparisons. For years it has been the general practice to use the LSD test to distinguish the true effects between any two means at a specified level of significance. However, it is not generally recognized that, in experiments involving more than two variables, the LSD test is applicable only to randomly selected or predetermined pairs of means.

It should be emphasized that this procedure is not appropriate when used indiscriminately to test all possible pairs of means. When the LSD test is applied in such a manner there is a likelihood of overestimating the significance of certain of the differences. The reason for this is that the difference between the highest and lowest mean in an array will on the average be greater than the difference between two randomly chosen means.

The use of the LSD test, therefore, calls for careful consideration in all cases where there are three or more means in the experiment and the F-test has demonstrated significant differences exist between them. It is not proposed to recommend complete abandonment of the use of the LSD test, but to limit its use to situations where the experimenter can make conclusions from his data with adequate decision-making protection. For other types of decision-making, another procedure will be discussed.

The following precautions and recommendations relate specifically to the use of the LSD test:

1) The LSD test should not be applied unless the F-test indicates that differences among means are present. A nonsignificant F-test should preclude any attempt to make further interpretations of differences among a set of means. The investigator should, under such circumstances, look upon widely separated means as worthy of future experimentation. Such mean differences cannot be interpreted on the basis of known probabilities.

2) The LSD test is valid only in testing mean comparisons which were identified in the objective of the experiment. This is a most important precaution since the testing of mean comparisons after the data are at hand introduces unknown probabilities and may lead to inaccurate conclusions. This subject will be discussed later in relation to the use of Duncan's Multiple Range Test.

3) The LSD test is reasonably satisfactory for comparing each of a set of means individually with a standard or a control.

4) The LSD test is reasonably satisfactory in determining the group of arrayed means (with an absence of noticeable gaps) that are not significantly different from the topmost mean. The plant breeder is frequently interested in selecting the high-yielding varieties or strains that are equal in yielding ability to the highest yielding individual in an experiment. These selections then form the group to be used for a subsequently more precise yield test.

Duncan's Multiple Range Test

In testing the significance of mean comparisons it is necessary that the testing procedure consider the probability of detecting real differences when they exist. It is generally preferred that the procedure assure the highest probability of rejecting the significance of mean differences when, in fact, the differences are not real. As indicated on page 3, this is considered protection against committing an error of the second kind (Type II error).

In addition to giving a high degree of inference protection, the procedure should be as simple of calculation as possible. Such a procedure is embodied in the Multiple Range Test proposed by Duncan (5), which employs special protection levels based on degrees of freedom. This test is similar to the range test proposed by Newman (12) and Keuls (10), but it uniformly assures the investigator of better protection against Type II errors.

It is apparent from the significant values in Table 3 that the LSD test is less conservative, in cases of more than two means, than the Duncan Multiple Range Test. Thus, with the LSD test, a "significant" value of 2.77 is uniformly required to detect significance between any two means in sets extending from 2 or more means in a range or group. Duncan's test is identical with the LSD test for two means but, on the other hand, requires progressively larger "significant values" as the number of means increase in a group.

Table 3. Comparisons of significant values at the 5-percent level of significance for various numbers of means, assuming a variance of unity and infinite degrees of freedom for experimental error

Kind of test procedure	Number of means in group					
	2	4	8	10	20	100
LSD	2.77	2.77	2.77	2.77	2.77	2.77
Duncan's Multiple Range	2.77	3.02	3.23	3.29	3.47	3.67
Newman-Keuls	2.77	3.63	4.29	4.47	5.01	5.93

Where a mean comparison test is appropriate, it is recommended that the research worker use Duncan's Multiple Range Test. The application of this range test will be illustrated in detail.

Analytical Procedure

The application of this multiple range test to a small set of means in a randomized complete block experiment will be given in detail. The variety means in Duncan's (6) paper will be used for this discussion. These, ranked in order from lowest to highest, are as follows:

A	F	G	D	C	B	E
49.6	58.1	61.0	61.5	67.6	71.2	71.3

The information required to apply this test to a set of data is: (1) the means; (2) the standard error of the mean ($s_{\bar{x}}$), derived from the error mean square, as

$$\sqrt{\frac{s^2}{n}},$$

where s^2 = the mean square for error and n = the number of replications; (3) the degrees of freedom (n_2) on which the mean square for error is based, and (4) a significant F-test. In this example, the seven varieties with means are as indicated above; a standard error of the mean, $s_{\bar{x}} = 3.643$; $n_2 = 30$ degrees of freedom for error in the analysis of variance; and a significant F-test.

The necessity for limiting the use of this multiple range test to cases where the F-test is significant has not been definitely established. However, in order to prevent contradictions between this range test and the F-test, it is recommended that this range test be used only when the F-test is significant.

In addition to the above information, special tables of "Significant Studentized Ranges" for the 5-percent and 1-percent levels of significance are needed. These are given in Tables 4 and 5.

The appropriate significant studentized ranges (r_p) for the various size of ranges (p) at the 5-percent level of significance, for this example, are taken from Table 4. The respective significant studentized ranges are each multiplied by the standard error of the mean to give what Duncan has termed the "shortest significant ranges" (R_p).

The above-mentioned shortest significant range values (R_p) for this example are:

p:	2	3	4	5	6	7
r_p:	2.89	3.04	3.12	3.20	3.25	3.29
R_p:	10.53	11.07	11.37	11.66	11.84	11.99

Table 4. Significant studentized ranges for the 5-percent level of significance [1/]

n_2 \ p	2	3	4	5	6	7	8	9	10	12	14	16	18	20	50	100
1	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0
2	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09
3	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50
4	3.93	4.01	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02
5	3.64	3.74	3.79	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83
6	3.46	3.58	3.64	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68
7	3.35	3,47	3.54	3.58	3.60	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61
8	3.26	3.39	3.47	3.52	3.55	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56
9	3.20	3.34	3.41	3.47	3.50	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52
10	3.15	3.30	3.37	3.43	3.46	3.47	3.47	3.47	3.47	3.47	3.47	3.47	3.47	3.48	3.48	3.48
11	3.11	3.27	3.35	3.39	3.43	3.44	3.45	3.46	3.46	3.46	3.46	3.46	3.47	3.48	3.48	3.48
12	3.08	3.23	3.33	3.36	3.40	3.42	3.44	3.44	3.46	3.46	3.46	3.46	3.47	3.48	3.48	3.48
13	3.06	3.21	3.30	3.35	3.38	3.41	3.42	3.44	3.45	3.45	3.46	3.46	3.47	3.47	3.47	3.47
14	3.03	3.18	3.27	3.33	3.37	3.39	3.41	3.42	3.44	3.45	3.46	3.46	3.47	3.47	3.47	3.47
15	3.01	3.16	3.25	3.31	3.36	3.38	3.40	3.42	3.43	3.44	3.45	3.46	3.47	3.47	3.47	3.47
16	3.00	3.15	3.23	3.30	3.34	3.37	3.39	3.41	3.43	3.44	3.45	3.46	3.47	3.47	3.47	3.47
17	2.98	3.13	3.22	3.28	3.33	3.36	3.38	3.40	3.42	3.44	3.45	3.46	3.47	3.47	3.47	3.47
18	2.97	3.12	3.21	3.27	3.32	3.35	3.37	3.39	3.41	3.43	3.45	3.46	3.47	3.47	3.47	3.47
19	2.96	3.11	3.19	3.26	3.31	3.35	3.37	3.39	3.41	3.43	3.44	3.46	3.47	3.47	3.47	3.47
20	2.95	3.10	3.18	3.25	3.30	3.34	3.36	3.38	3.40	3.43	3.44	3.46	3.46	3.47	3.47	3.47
22	2.93	3.08	3.17	3.24	3.29	3.32	3.35	3.37	3.39	3.42	3.44	3.45	3.46	3.47	3.47	3.47
24	2.92	3.07	3.15	3.22	3.28	3.31	3.34	3.37	3.38	3.41	3.44	3.45	3.46	3.47	3.47	3.47
26	2.91	3.06	3.14	3.21	3.27	3.30	3.34	3.36	3.38	3.41	3.43	3.45	3.46	3.47	3.47	3.47
28	2.90	3.04	3.13	3.20	3.26	3.30	3.33	3.35	3.37	3.40	3.43	3.45	3.46	3.47	3.47	3.47
30	2.89	3.04	3.12	3.20	3.25	3.29	3.32	3.35	3.37	3.40	3.43	3.44	3.46	3.47	3.47	3.47
40	2.86	3.01	3.10	3.17	3.22	3.27	3.30	3.33	3.35	3.39	3.42	3.44	3.46	3.47	3.47	3.47
60	2.83	2.98	3.08	3.14	3.20	3.24	3.28	3.31	3.33	3.37	3.40	3.43	3.45	3.47	3.48	3.48
100	2.80	2.95	3.05	3.12	3.18	3.22	3.26	3.29	3.32	3.36	3.40	3.42	3.45	3.47	3.53	3.53
∞	2.77	2.92	3.02	3.09	3.15	3.19	3.23	3.26	3.29	3.34	3.38	3.41	3.44	3.47	3.61	3.67

1/ Acknowledgment is given to Dr. David B. Duncan and Professor Gertrude M. Cox, Editor of Biometrics, for permission to publish this table.

Table 5. Significant studentized ranges for the 1-percent level of significance [1/]

n_2 \ p	2	3	4	5	6	7	8	9	10	12	14	16	18	20	50	100
1	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0
2	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0
3	8.26	8.5	8.6	8.7	8.8	8.9	8.9	9.0	9.0	9.0	9.1	9.2	9.3	9.3	9.3	9.3
4	6.51	6.8	6.9	7.0	7.1	7.1	7.2	7.2	7.3	7.3	7.4	7.4	7.5	7.5	7.5	7.5
5	5.70	5.96	6.11	6.18	6.26	6.33	6.40	6.44	6.5	6.6	6.6	6.7	6.7	6.8	6.8	6.8
6	5.24	5.51	5.65	5.73	5.81	5.88	5.95	6.00	6.0	6.1	6.2	6.2	6.3	6.3	6.3	6.3
7	4.95	5.22	5.37	5.45	5.53	5.61	5.69	5.73	5.8	5.8	5.9	5.9	6.0	6.0	6.0	6.0
8	4.74	5.00	5.14	5.23	5.32	5.40	5.47	5.51	5.5	5.6	5.7	5.7	5.8	5.8	5.8	5.8
9	4.60	4.86	4.99	5.08	5.17	5.25	5.32	5.36	5.4	5.5	5.5	5.6	5.7	5.7	5.7	5.7
10	4.48	4.73	4.88	4.96	5.06	5.13	5.20	5.24	5.28	5.36	5.42	5.48	5.54	5.55	5.55	5.55
11	4.39	4.63	4.77	4.86	4.94	5.01	5.06	5.12	5.15	5.24	5.28	5.34	5.38	5.39	5.39	5.39
12	4.32	4.55	4.68	4.76	4.84	4.92	4.96	5.02	5.07	5.13	5.17	5.22	5.24	5.26	5.26	5.26
13	4.26	4.48	4.62	4.69	4.74	4.84	4.88	4.94	4.98	5.04	5.08	5.13	5.14	5.15	5.15	5.15
14	4.21	4.42	4.55	4.63	4.70	4.78	4.83	4.87	4.91	4.96	5.00	5.04	5.06	5.07	5.07	5.07
15	4.17	4.37	4.50	4.58	4.64	4.72	4.77	4.81	4.84	4.90	4.94	4.97	4.99	5.00	5.00	5.00
16	4.13	4.34	4.45	4.54	4.60	4.67	4.72	4.76	4.79	4.84	4.88	4.91	4.93	4.94	4.94	4.94
17	4.10	4.30	4.41	4.50	4.56	4.63	4.68	4.72	4.75	4.80	4.83	4.86	4.88	4.89	4.89	4.89
18	4.07	4.27	4.38	4.46	4.53	4.59	4.64	4.68	4.71	4.76	4.79	4.82	4.84	4.85	4.85	4.85
19	4.05	4.24	4.35	4.43	4.50	4.56	4.61	4.64	4.67	4.72	4.76	4.79	4.81	4.82	4.82	4.82
20	4.02	4.22	4.33	4.40	4.47	4.53	4.58	4.61	4.65	4.69	4.73	4.76	4.78	4.79	4.79	4.79
22	3.99	4.17	4.28	4.36	4.42	4.48	4.53	4.57	4.60	4.65	4.68	4.71	4.74	4.75	4.75	4.75
24	3.96	4.14	4.24	4.33	4.39	4.44	4.49	4.53	4.57	4.62	4.64	4.67	4.70	4.72	4.74	4.74
26	3.93	4.11	4.21	4.30	4.36	4.41	4.46	4.50	4.53	4.58	4.62	4.65	4.67	4.69	4.73	4.73
28	3.91	4.08	4.18	4.28	4.34	4.39	4.43	4.47	4.51	4.56	4.60	4.62	4.65	4.67	4.72	4.72
30	3.89	4.06	4.16	4.22	4.32	4.36	4.41	4.45	4.48	4.54	4.58	4.61	4.63	4.65	4.71	4.71
40	3.82	3.99	4.10	4.17	4.24	4.30	4.34	4.37	4.41	4.46	4.51	4.54	4.57	4.59	4.69	4.69
60	3.76	3.92	4.03	4.12	4.17	4.23	4.27	4.31	4.34	4.39	4.44	4.47	4.50	4.53	4.66	4.66
100	3.71	3.86	3.98	4.06	4.11	4.17	4.21	4.25	4.29	4.35	4.38	4.42	4.45	4.48	4.64	4.65
∞	3.64	3.80	3.90	3.98	4.04	4.09	4.14	4.17	4.20	4.26	4.31	4.34	4.38	4.41	4.60	4.68

1/ Acknowledgment is given to Dr. David B. Duncan and Professor Gertrude M. Cox, Editor of Biometrics, for permission to publish this table.

The various values for r_p are found on the line for 30 degrees of freedom in Table 4 at the respective values of p (p = number of varieties in the range). The shortest significant range for each value of p, as indicated above, is obtained by multiplying each by the standard error of the mean (3.643). Thus, the value of R_p = 2 is 2.89 x 3.643 = 10.53.

The variety mean differences are contrasted with the shortest significant range appropriate to the range (p) under test in the following order: the largest minus the smallest, the largest minus the second smallest, up to the largest minus the second largest; then the second largest minus the smallest, the second largest minus the second smallest, and so on up to the second smallest minus the smallest. Applying the above procedure to all possible combinations of mean differences in this example, we get

E-A (p=7) = 71.3-49.6 = 21.7*
E-F (p=6) = 71.3-58.1 = 13.2*
E-G (p=5) = 71.3-61.0 = 10.3
E-D (p=4) = 71.3-61.5 = 9.8
E-C (p=3) = 71.3-67.6 = 3.7
E-B (p=2) = 71.3-71.2 = 0.1
B-A (p=6) = 71.2-49.6 = 21.6*
B-F (p=5) = 71.2-58.1 = 13.1*
B-G (p=4) = 71.2-61.0 = 10.2
B-D (p=3) = 71.2-61.5 = 9.7
B-C (p=2) = 71.2-67.6 = 3.6

C-A (p=5) = 67.6-49.6 = 18.0*
C-F (p=4) = 67.6-58.1 = 9.5
C-G (p=3) = 67.6-61.0 = 6.6
C-D (p=2) = 67.6-61.5 = 6.1
D-A (p=4) = 61.5-49.6 = 11.9*
D-F (p=3) = 61.5-58.1 = 3.4
D-G (p=2) = 61.5-61.0 = 0.5
G-A (p=3) = 61.0-49.6 = 11.4*
G-F (p=2) = 61.0-58.1 = 2.9
F-A (p=2) = 58.1-49.6 = 8.5

The asterisks following some of the mean differences above indicate that those differences exceed the value of the shortest significant range for the respective values of the significant studentized ranges (r_p) at the 5-percent level of significance. Thus, the mean difference of 21.7 for E-A, which represents a studentized significant range of p = 7, exceeds the value of the shortest significant range of 11.99. Likewise, the mean difference for E-F of 13.2 exceeds the value of the shortest significant range of 11.84, for a range of p = 6. On the other hand, the value of F-G = 10.3 does not exceed the shortest significant range of 11.66 for p = 5. Therefore, the remainder of the mean differences involving variety E are likewise nonsignificant.

The rule for the above procedure, with one exception, as given by Duncan (5, p.6) is as follows:

> "Each difference is significant if it exceeds the corresponding shortest significant range; otherwise, it is not significant."

> The exception rule being "that no difference between two means can be declared significant if the two means concerned are both contained in a subset of the means which has a non-significant range."

As a result of the above rule and exception, it is convenient to group the two means found to be nonsignificantly different with all of the intervening means and underscore them.

Step 1. Taking the array of means comprising this example, we would underscore the varieties G, D, C, B, and E, as done in diagram (1) below,

49.6	58.1	61.0	61.5	67.6	71.2	71.3
A	F	G	D	C	B	E ...(1)

Thus, applying the exception rule, the remaining differences between all varieties of the underscored subset are nonsignificant. As a result, the intervening means need not be tested against the mean significant differences. As a consequence and since E-G is nonsignificant, the following mean comparisons are also nonsignificant: E-D, E-C, E-B; B-G, B-D, B-C; C-G, C-D; and D-G.

Step 2. Starting with the second highest yielding variety (B), it is clear from the underscored varieties shown above that B-A and B-F are both significant. No line need be added to indicate this relationship since it is already apparent from the underscoring of G, D, C, B, E. The other comparisons contrasted with variety B (B-G, B-D, and B-C) are nonsignificant as demonstrated in Step 1.

Step 3. Comparisons are now made with variety C and all varieties of lower yielding ability. It is clear from the tabulation of mean differences above that C-A is significant and C-F is not. Applying the exception rule, comparisons C-G, C-D; D-F, D-G; and G-F are not significant. These results are indicated by a second underscoring to include F, G, D, C, as

49.6	58.1	61.0	61.5	67.6	71.2	71.3
A	F	G	D	C	B	E ...(2)

Step 4. Using variety D for comparing with those varieties of lower yielding ability and the above-mentioned tabulation (p. 27) of mean differences, it is evident that D-A is significant, whereas D-F is not significant from Step 3 and D-G is nonsignificant from Steps 1 or 3.

Step 5. Comparing varieties A and F with G, we find that G-A is significant, but G-F is nonsignificant from Step 3. No addition line is needed to indicate this since these two relationships are included in diagram (2).

Step 6. The remaining comparison of F-A is nonsignificant (see

tabulated mean differences on page 27). Since these two means comprise a subset of means with a nonsignificant range (exception rule) an underscore is made between A and F as indicated in diagram (3) below:

49.6	58.1	61.0	61.5	67.6	71.2	71.3	
A	F	G	D	C	B	E	...(3)

With a little practice each of the steps can be carried out almost by inspection.

Data Presentation in Tables

The relative newness of Duncan's Range Test presents a problem to the investigator as to the method of presentation of experimental data in tabular form. Designation of significance between mean differences in a table may be indicated in two ways, e.g., (1) by underscoring means comprising a group that are not significantly different or (2) by the use of letters to differentiate significantly different means.

The underscoring procedure for two different types of tables is illustrated in Tables 6 and 7. The differentiation of means with letter designations is presented in Table 8.

The application of the underscoring procedure to a set of data from a 2x6 factorial experiment is illustrated in Table 9. In dealing with this type of data, the overall means for cover crops are ranked in ascending order of magnitude. It will be noted that the departure from ranked order (nonsignificant interaction) within the plowing dates did not affect the underscoring procedure for the identification of the homogeneous groups. Situations where significant interactions exist could be handled by the method presented in Table 8.

Table 6. Average percentage bunt in 9 regional tests of seed treatment fungicides for the control of bunt in winter wheat[1]

Treatment	Rate (oz. per bushel)	Percentage bunt	Statistical significance
Check - untreated		55.3	
Panogen 15	0.75	34.4	
Panogen 15	1.5	26.3	
Capton - HCB 50-20	0.5	20.9	
Terraclor	0.5	11.1	
Dow HCB	0.5	10.3	
Capton - HCB 50-20	1.0	9.6	
No Bunt	0.5	8.3	
Sanocide	0.5	7.7	
Terraclor	1.0	6.5	
Anticarie	0.5	5.6	
Capton - HCB 50-20	2.0	5.5	
No Bunt	1.0	4.6	
Dow HCB	1.0	4.6	
Panogen 15	3.0	3.8	
Anticarie	1.0	3.3	
Dow HCB	2.0	2.9	
Terraclor	2.0	2.9	
Sanocide	1.0	2.4	
Sanocide	2.0	2.1	
Anticarie	2.0	2.0	
No Bunt	2.0	2.0	

[1] Table 2 of article by Laurence H. Purdy, published in U.S.D.A., Plant Disease Reporter 39:844-849. 1955

Table 7. Average height of fall growth of lots of Ranger alfalfa produced at different locations[1]

	Average height of plants (inches) grown from seed produced at specified locations.						
	Mont. F	Wash. R	Mont. C	Calif. C	Ariz. C	Mex. C	Ariz.
Height, inches	3.4	3.5	3.6	3.8	4.1	4.9	5.3
Statistical significance							

[1] Modified from Table 1 of article by Dale Smith, Agron. Jour. 47: 201-205. 1955

Table 8. Effect of location on the oil content of corn grain.[1]

Location	Year	Mean (%)	Stat. Sig.[2]
Williamsburg	1948	4.4	a
Petersburg	1948	4.3	b
Charlotte Courthouse	1949	4.2	c d
Staunton	1949	4.2	c d
Onley	1948	4.1	d
Chatham	1949	4.0	e
Blacksburg	1949	3.9	e

[1] Part of Table 1 of article by C. F. Genter, J. F. Eheart, and W. N. Linkous, published in Agron. Jour. 48:63-70. 1956

[2] Stat. Sig. means statistical significance. Mean followed by letter "a" is significantly different from those means not having "a"; those followed by "b" are significantly different from those not having "b", etc.

Table 9. Yields (in lbs.) of tobacco as influenced by time of plowing under different cover crops[1]

Kind of cover crop	Time of Plowing				Cover crop Average	
	Early		Late			
	Yield	Stat. Sig.[2]	Yield	Stat. Sig.[2]	Yield	Stat. Sig.[2]
None	940		932		936	
Ryegrass	1013		955		984	
Ryegrass-vetch	1226		1396		1311	
Vetch	1219		1427		1323	
Rye-vetch	1290		1500		1395	
Wheat-vetch	1343		1457		1400	

[1] Data reported by C. S. Britt and C. S. Slater at the 1956 Meeting of the American Society of Agronomy.

[2] Stat. Sig. means statistical significance.

LITERATURE CITED

(1) Anderson, R. L., and E. E. Houseman
1942 Tables of orthogonal polynomial values extended to N=104. Ia. Agr. Exp. Sta. Res. Bul. 297:595-672.

(2) Cochran, William G., and Gertrude M. Cox
1950 Experimental designs. John Wiley and Co., New York, 454pp.

(3) Duncan, D. B.
1951 A significance test for differences between ranked treatments in an analysis of variance. Va. Jour. Sci. 2:171-189.

(4) _____
1952 On the properties of the multiple comparisons test. Va. Jour. Sci. 3:49-67.

(5) _____
1955 Multiple range and multiple F tests. Biometrics 11:1-42.

(6) Federer, Walter T.
1955 Experimental design. Theory and Application. Macmillan Company, New York. 544 pp.

(7) Fisher, R. A.
1935 The design of experiments. Oliver and Boyd, London. 6th Ed., 244 pp., (Section 24).

(8) _____
1954 Statistical methods for research workers. Hafner Pub. Co. Inc., New York. 356 pp.

(9) _____ and Frank Yates
1948 Statistical tables for biological agricultural and medical research. Hafner Pub. Co. Inc., New York. 112 pp.

(10) Keuls, M.
1952 The use of the "Studentized range" in connection with an analysis of variance. Euphytica 1:112-122.

(11) Larmour, R. K.
1941 A comparison of hard red spring and hard red winter wheats. Cereal Chem. 18:778-789.

(12) Newman, D.
1939 The distribution of range in samples from a normal population, expressed in terms of an independent estimate of standard deviation. Biometrika 31:20-30.

(13) Scheffe, Henry
1953 A method for judging all contrasts in the analysis of variance. Biometrika 40:87-104.

(14) Tukey, J. W.
1951 Quick and dirty methods in statistics. Part II, Simple analyses for standard designs. Proc. 5th Ann. Conven. Amer. Soc. Qual. Control. p-189-197.

(15) _____
1953 The problem of multiple comparisons. [Unpublished notes] Princeton University, 396 pp.